Data Sharing: Low-Cost Sensors for Affect and Cognition

Keith Brawner
United States Army Research Laboratory
12423 Research Parkway
Orlando, Florida, 32826
Keith.W.Brawner@us.army.mil

ABSTRACT

The Educational Data Mining (EDM) community has experienced many benefits from the open sharing of data. Efforts such as the Pittsburgh Science of Learning Center Datashop have helped in the development of learning data storage and standards in the educational community. In other fields, standards of comparison have been created through publication, sharing, and competition on identical datasets. This ability to share, compare, and grow as a field has proven to be a success point for the community of scientists. This paper presents a new and unique dataset intended for sharing with the educational data mining community which is unique in its features, measures affective and cognitive elements into learning experiences, and presents a challenge for EDM researchers. Initial offline analysis results and secondary online analysis results are presented as benchmarks for comparison by future researchers.

Keywords

Data mining, machine learning, data sharing, affect, cognition

1. INTRODUCTION

The most influential example of data sharing in the educational data mining field is doubtless the Pittsburgh Science of Learning Center Datashop [26]. The PSLC DataShop presents fine-grained, extensive, and longitudinal data on student learning, including demographic information on students. At the time of writing, the PLSC DataShop has been cited within over 100 pieces of academic writing, where numerous researchers have benefited. This process of data sharing is pushed academically both from the National Science Foundation [4], and from the Army Research Laboratory (ARL) through the National Academy of Sciences [1].

The United States Army Research Laboratory (ARL) has a mission of "Technology Driven. Warfighter focused", and occupies an unusual space in the world of scientific progress. Most academic research institutions have an emphasis on discovery, invention, innovation, and collaborative sharing. Most commercial institutions have an emphasis on providing market-driven solutions to customer problems. Between these two areas lies the problem of transition, where innovations and inventions find application within an area of customer use [17]. ARL has a primary interest in creating original research in areas of defense interest, and attempting to accelerate the both the pace of discovery through collaboration with research institutions, and the rate of technology adoption, through collaboration with a user community composed of both industrial and military technology adopters. Efforts to focus research effort upon user-relevant problems occupy the space between research and commercialization, and consequently offer an attractive manner to speed both. The open sharing of research data presents such an option, both through aiding academic researchers and through the transition of the subsequently developed research solutions into a field of use.

The author of this paper intends to share a unique dataset as part of the submission of this paper, discuss the reasons for its collection, as well as the experiment which produced it. Additionally, the author presents some initial analysis showing methods for classification of this data can be developed, and some of the experiments in producing rapid classifiers which indicate a need for feature extraction. Each of these experiments can benefit from additional research experimentation, as well as a commercialization testbed. These initial analyses provide baselines benchmarks for comparison by future researchers. Finally, there is some discussion of additional datasets which may be able to be shared in the near future.

2. DATA SHARING

The sharing of data allows researchers to compare research results, be assured that they are working on a similar problem, and presents an avenue for collaboration. Open competition and cooperation among data mining scientists has resulted in numerous improvements to overall solution quality [12]. Oftentimes these solutions are the result of incremental progress among a community, or are spurred on as a result of the competitive nature of research or commercialization. Data sharing may have one or more of the following primary goals:

- To develop standards, as part of sequential data processing techniques
- To improve the overall state of practice, as part of benchmarking
- To give share in the efforts of other researchers, including the data collection, cleanup, feature extraction, classification algorithms, and other items.

However, efforts in data sharing additionally benefit from having some idea of the original project and motivations. Knowledge of the initial motivation and initial data analysis helps to explain the purpose of what is being looked for, as well as provide a starting place when pursuing individual research interests. An initial description of the study which produced the data, with information on locations for where to get further and in-depth descriptions, significantly aids the community ability to mine the data for knowledge. This has become the standard of practice in many data mining communities [12; 26].

3. INITIAL MOTIVATION

The ARL Learning in Intelligent Tutoring Environments (LITE) Lab has an interest in Intelligent Tutoring Systems (ITS) research, and has developed the Generalized Intelligent Framework for Tutoring (GIFT) [32] as an architectural output for research. GIFT is composed of several interoperable modules for the communication of sensor, learner, instructional, and performance information. As a part of ongoing GIFT research, there are various projects examining the state of the art for each of these areas. Among the goals of the GIFT project is to be able to rapidly transition performed research into the open-source community. Transition tools, authoring tools, and multiple programming language plugins have been constructed for this purpose, are curated to ensure overall stability and use, and are freely and publicly distributed [3].

The Sensor Module component of GIFT is tasked with the measurement and communication of sensor information to the remaining components of the architecture. The remaining components of the architecture, in turn, transform this sensor information into pedagogical strategies designed to influence learning gains [18]. In an ideal world, these functions would be performed with perfect accuracy, no cost, and in an unobtrusive fashion. Currently, however, sensors have a tendency to be intrusive, high cost, and only somewhat reliable. The measurement of affective and cognitive states for educational purposes is of interest to the community, and steps to do so cheaply with little human interaction are desirable.

These affective phenomenon (e.g. Anger, Boredom, Anxiety, etc.) have been shown to have impact over learning outcomes [31], which make it desirable to measure them as part of the interaction with an educational system [33]. As an example, high arousal relates to increased attention to the arousing item [25], boredom leads to lower retention and information application [31], and frustration can distract from learning activities [30]. Classification of these states is of interest to the GIFT architecture, as it is able to recommend variation in pedagogical intervention based upon their availability.

Cognitive phenomenon (e.g. attention, working memory, executive function, etc.) may also be of interest to the creators of learning systems. As an example, divided attention leads to lower learning performance [31], low engagement may indicate non-participation within a learning environment [16], and high workloads show a decrease in both performance and retention [20]. A system which is able to measure, predict, or model the emergence of both cognitive and affective phenomenon may be able to increase learning outcomes, which is a significant part of the goal of the educational data mining community and LITE Lab alike.

Currently, the measurement of these states is dependent upon the means of detecting them. These states are detected via bodily sensors or via self-reporting data. The set up of bodily sensors, or the query of self-report information takes time away from learning activities, can be uncomfortable, can disengage from the learning environment, and can be expensive to purchase. It is desirable to be able to detect these states with little interference, and at little cost. The initial motivation for the collection of the subject dataset was to reduce the cost of sensor-based methods of learning state detection, as has been published as part of this goal [13; 27]. The sensors used during this experiment were extremely low cost, with less than $300 spent on the total sensor suite, when excluding the sensors used to produce labeling information for participants. If algorithmic classification for these low cost sensors is able to mimic (albeit with lessened accuracy) the performance of the high cost sensors, then it creates the ability for tradeoffs between cost and resolution for the creator of intelligent tutoring systems.

4. EXPERIMENT

Participants were taken from a pool of General Psychology Students at the United States Military Academy (USMA) at Westpoint. Participants took place in two parts of a study in order to induce various affective and cognitive states. During the first phase, participants were asked to undertake a visual vigilance task, watch video clips from the movie Halloween, and the movie My Bodyguard. The video segment from Halloween has been previously validated to induce Fear and Anxiety, while the video segment from My Bodyguard has previously been validated to induce Anger and Frustration [22].

During the second phase of the study, participants were asked to take place in several scenarios within the Army's Virtual Battlespace 2 (VBS2) video game. VBS2 is used for training a variety of Army-relevant tasks. Four VBS2 scenarios which had previously been validated to produce various cognitive and affected states were played [24]. The primary affective or cognitive inducing event experienced, in each scenario, was:

1. Limited visual perceptions within a tactical and dangerous scenario environment
2. An unrealistically large number of enemies
3. Annoying sounds
4. Equipment malfunction

The cognitive and affective states, and the tasks which induced them, are presented in Table 1. The author maintains that the purpose of this paper is to share the data and analyses, rather than to present the total research study. The study is described in greater depth in other publications [8; 13; 27].

Table 1. Summary of induced affective and cognitive states

Affective State			
	Boredom	Anxiety / Fear	Anger / Frustration
Task	Visual vigilance		
Movie Clip		Halloween	My Bodyguard
VBS2 Scenario		34	1234
Cognitive State			
	Workload	Engagement	Distraction
VBS2 Scenario	1234	1234	1234

4.1 Hardware

Figure 1 shows a fully instrumented participant. In total, measurements were collected via two Electroencephalography (EEG) systems (from Neurosky and Advanced Brain Monitoring (ABM)), a custom-made eye tracker, a Zephyr heart rate monitor, embedded Phidget pressure sensors within the chair, a Venier sonar sensor for distance from the computer, and emotional self-report measure. The self-report measure of EmoPro and the ABM headset have previously been validated to produce accurate measures of affective and cognitive states, respectively [14; 23]. A summary of the measures which these sensors produce is included in Table 2. Larger discussion on the relevance of each of these states to learning outcomes and validation of the baseline measurements is available in prior work [8; 13; 27].

Figure 1. Fully Instrumented Participant

Table 2. Summary of sensors used and states measured.

Sensor	Affective State	Cognitive State
*ABM EEG**		Attention, Engagement, Distraction, Drowsiness, Workload
Neurosky EEG		
Eye-tracker		Attention, Drowsiness, Workload
*EmoPro**	Anger, Anxiety, Arousal, Boredom, Fear, Stress	Attention
Heart Rate Monitor		
Chair Pressure Sensor (posture)	Arousal, Boredom, Frustration	Engagement, Flow
Motion Detector (posture)		
** Indicates Ground Truth Measurement*		

5. INITIAL ANALYSIS

The Logistic Model Tree (LMT) method of analysis [28] was selected for classifier construction on this data from among a series of methods considered [27]. Ten-fold cross validation was used in an effort to avoid overfitting. The created trees were found to have a single node, rendering this method similar to logistic regression. The measure of Area Under the Curve (AUC) of the Receiver Operator Characteristic (ROC) [21] is used to evaluate overall model quality. In general, the AUC ROC method produces a value in the range [0,1], with 1.0 representing perfect classification accuracy and 0.5 representing baseline levels. The overall finding is that there is significant room for improvement of generalized model quality, but that data trends are available to do so. These findings are summarized with Table 3 and Table 4.

Table 3. Summary of which sensor data was used to create Initial Emotional Models

Low-Cost Sensor	EmoPro Measurements		
	Anger	Anxiety/Fear	Boredom
HR			X
Eye Track			
EEG		X	X
Chair		X	
Distance		X	X
AUC ROC	**N/A**	**0.83**	**0.79**

Table 4. Summary of which sensor data was used to create Initial Cognitive Models

Low-Cost Sensor	ABM Measurements		
	Engagement	Distraction	Workload
HR	X	X	
Eye Track			
EEG			
Chair	X	X	X
Distance	X		X
AUC ROC	**0.80**	**0.81**	**0.82**

6. REPURPOSED ANALYSIS

Later projects were investigating a realtime signal approach to data processing for the classification of emotional states in realtime [10]. The core idea of this approach is that highly adaptable and individualized approaches to modeling would be better able to model emerging states. There is some evidence that adaptable approaches among cognitive state data are able to model more accurately, but there remain few attempts to model states in this way [2]. Additionally, there is evidence that models created from bodily sensor data may fail generalization tests for reasons such as electrode drift, changes in default impedance, and other non-linear behavioral factors [2].

In order to perform this type of modeling, there needs to be a dataset with a number of desirable features. It needs to include states of relevant interest to learning, and the ability to transfer these states to other systems (sensor-based data streams). It must be created on a relevant population (learners), with items which can be potentially included within classrooms. Finally, work for this purpose must contain high resolution labeling information, and contain previously established models for the purpose of comparison. Having all of these features present in a single dataset is rare, but present in the dataset described by this paper. This checklist is presented in Table 5, as a checklist, in an effort to recommend similar types of data collection for other experimental designers.

Table 5. Checklist of features for Low Cost Sensor dataset (recommended for other studies)

Does the dataset have…	?
Relevant states to learning	
Ability to be transferred, without modification, to another domain of instruction	
Relevant population	
Relevant cost for classroom inclusion	
Labeled data	
Initial benchmarks for research comparison	

6.1 Secondary Analysis

The author sought a dataset with relevant features (Table 5) for future work in testing various schemes of classification, for simultaneous model creation and use within a training environment. While simultaneous model creation and use allows for a highly individualized approach, it requires special treatment of the datastream. In brief, any appropriate method must treat the stream as though it has infinite length, be able to discover emergent concepts, keep track of these concepts as they are redefined, and separate them from other distinct areas. These fundamental problems are discussed at length in order work, but importantly preclude the application of traditional methods such as reinforcement learning, Bayesian, or genetic approaches, primarily because of the infinite length problem [5; 8].

The approaches of Adaptive Resonance Theory (ART) [11], Online Semi-Supervised Growing Neural Gas (GNG) [6], Vowpal Wabbit (VW) [29], and incremental clustering [9] were selected to create, use, and evaluate a model simultaneously [8]. Each of these methods creates a class, cluster, or other area of labeled information around an area of the sampling space in various manners. As an example, clustering creates new classes based on a parameter-specified distance away from an existing cluster centroid and classifies a point to cluster membership based on the nearest cluster centroid while ART creates new classes and classifies classes based upon representative vectors.

Each of the four methods mentioned above was modified to create a semi-supervised learning approach which required a significantly lower number of labeled data points in order to make classification decisions. This type of approach allows for limited labeling information to exploit cluster/class/group structure and search efficiently through a set of hypothesis derived among mostly-unlabeled information [15]. This was further engineered to take an active learning approach, where each algorithm was enabled to select the points which appeared most helpful towards aiding these classification decisions. This type of approach presupposes the presence of an ‘oracle’ which is capable of granting label requests infrequently, which is a reasonable assumption when the student is available for questioning.

The total of these efforts is the development of realtime algorithmic approaches which are able to classify with very little labeling information. These approaches can be compared side-by-side to the binary classification, regression-based, logistic model trees created in the earlier study. Using methods for individualized realtime model construction on multiple individuals provides evidence to how well the model is likely to transfer to a new population, while having a comparison benchmark assures that it is possible to create a model at all. While the offline models used in the initial analysis are created with “infinite” time and 100% of the data, the offline models have limited time and data availability.

A variety of measures and feature elimination were used in the analysis, but it appears that something more rich needs to be performed in order to achieve parity performance with the offline models [8]. The measures included were “how well does this algorithm model the previous 10% of the total datastream?”, “how well does this algorithm predict the next 10% of the total datastream?”, and “how well does this algorithm model all of the data seen thus far?” An example chart of performance shows average model quality over the previous 100%, previous 10%, and next 10% averaged over all participants is shown in Figure 2 for the GNG algorithm. Figure 2 shows the easiest, supervised, setting using the GNG method, while Figure 3 shows performance of the ART method the most demanding, unsupervised, setting. Semi-supervised and active learning settings were found to have the expected performance in between the two differing levels of supervision. However, no method was able to construct quality cognitive models using the identical data used by the offline linear regression models.

The fundamental question that this research sought to answer, with regard to cognitive models, was “Can a quality model of cognition be constructed on an individualized and realtime basis, with various state-of-the-art methods, while using only the data which is used for the offline regression models?”. The answer, unfortunately, is that they cannot be constructed out of the raw data, due to the rapidly changing nature of cognitive states. Researchers who are not interested in confining themselves to a limited subset of datastreams, and have the freedom to feature engineer and reprocess as interested may well be able to create more usable cognitive models.

While the cognitive modeling conclusions are disappointing, the affective components of this secondary analysis show promise. In short, affective states are slow-changing enough that rapid model development shows promise for correlation with other observed data to represent finite and discriminable states. This work has been performed [9], but deserves more than the glancing discussion of algorithms, adaption, semi-supervised learning, and active learning which is presented within this data sharing paper. The affective work is currently under review for journal publication.

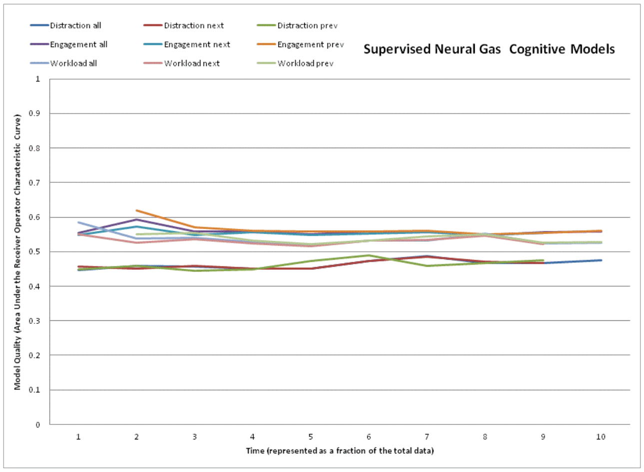

Figure 2. Subset of performance of cognitive models, using various metrics of viability. Subset shown is for supervised classification, using GNG method. Baseline measurements of 0.6 or greater are not achieved.

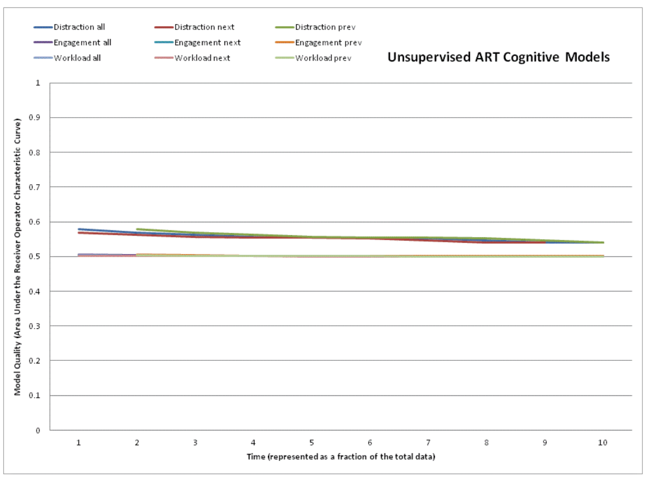

Figure 3. Subset of performance of cognitive models, using various metrics of viability. Subset shown is for unsupervised classification, using ART method. Baseline measurements of 0.6 or greater are not achieved.

7. CONCLUSION

The dataset in this paper has been collected at expense to the Army, but is useful to a wider public. An initial project analyzed this dataset in order to determine if low cost sensors are able to mimic the performance of high cost sensors when supplemented with classification improvements. The finding was that they were able to, but that more work was needed in order to mimic the performance of the high cost sensors in a generalized fashion with the data available.

A secondary look at this dataset investigated a different research question. This study sought to examine whether highly individualized (not generalized) cognitive models could be created with the same data available to previous classifiers. The answer to this question was that it could not be done with the raw cognitive data alone, and that further work would need to be done to develop filters, feature extraction, and other, differing, methods of processing.

The findings of both of these studies indicate that there is significant work left to be performed on this dataset. This includes the development of feature extraction techniques for more advanced model development, the development of additional types of classification, the development of online and realtime classifiers, and other items. This data could, of course, be used in new and unexpected ways by other researchers for purposes like cognitive and affective state correlation or developing data fusion approaches.

The sharing of this data with the community gives a unique opportunity for educational, or other, data mining researchers to analyze a set of data without the trouble of collection. This data is shown to be relevant to learning, collected with low cost sensors, within real world (noisy) conditions. A researcher who analyzes this data has the opportunity to expand directly on work performed across several communities of interest. These communities include the educational data mining, physiological sensor, EEG, military application, and Intelligent Tutoring System (ITS) areas of research, where portions of this research have before been published.

8. FUTURE WORK

The author encourages others to open datasets for joint sharing and analysis in an effort to close the loop between educational data research and an area of practice. An effort led by Art Graesser at this year's Educational Data Mining conference is focused on the joint benefits to community development while increasing the overall rate of experimentation and technology transition.

In addition to sharing the Low Cost Sensor Dataset, there are several other datasets which have been collected and face the possibility for sharing after internal hurdles are cleared. These include sets of GSR and ECG data [7], low-cost EEG data [19], and a currently unpublished study using the Microsoft Kinect®. As a part of the scientific community, it is the intention of the author, on behalf of ARL, to open up future datasets for re-analysis, future work and other forms of collaboration.

9. ACKNOWLEDGEMENTS

The researchers of this paper would like to acknowledge the large number of collaborators on this effort. We thank USMA for supporting data collection efforts, Dr. Mike Matthews and Kokini, Christina for support of the experiment, Ruben Ramirez-Padron for the initial analysis and benschmark comparison of the dataset. This research is supported as part of the Learning in Intelligent Tutoring Environments (LITE) Laboratory at ARL.

10. REFERENCES

[1] 2014. *2013-2014 Assessment of the Army Research Laboratory: Interim Report*. The National Academies Press.

[2] ALZOUBI, O., CALVO, R., and STEVENS, R., 2009. Classification of EEG for Affect Recognition: An Adaptive Approach. *AI 2009: Advances in Artificial Intelligence*, 52-61.

[3] ARL, 2012. Generalized Intelligent Framework for Tutoring Release Page Army Research Laboratories, http://www.gifttutoring.org.

[4] ATKINS, D., 2003. Revolutionizing science and engineering through cyberinfrastructure: Report of the

National Science Foundation blue-ribbon advisory panel on cyberinfrastructure.

[5] BERINGER, J. and HÜLLERMEIER, E., 2006. Online clustering of parallel data streams. *Data Knowl. Eng 58*, 2, 180-204.

[6] BEYER, O. and CIMIANO, P., 2011. Online semi-supervised growing neural gas. In *Workshop New Challenges in Neural Computation 2011*, 21.

[7] BRAWNER, K. and GOLDBERG, B., 2012. Real-Time Monitoring of ECG and GSR Signals during Computer-Based Training Springer, 72-77.

[8] BRAWNER, K.W., 2013. Modeling Learner Mood In Realtime Through Biosensors For Intelligent Tutoring Improvements. In *Department of Electrical Engineering and Computer Science* University of Central Florida, 500.

[9] BRAWNER, K.W. and GONZALEZ, A.J., 2011. Realtime Clustering of Unlabeled Sensory Data for User State Assessment. In *Proceedings of the Proceedings of International Defense & Homeland Security Simulation Workshop of the I3M Conference* (Rome, Italy, September 2011).

[10] BRAWNER, K.W., SOTTILARE, R., and GONZALEZ, A., 2012. Semi-Supervised classification of realtime physiological sensor datastreams for student affect assessment in intelligent tutoring. In *Intelligent Tutoring Systems* Springer, 582-584.

[11] CARPENTER, G.A. and GROSSBERG, S., 1995. Adaptive resonance theory (ART). In *The handbook of brain theory and neural networks*, M. ARBIB Ed. MIT press, Cambridge, MA, 79-82.

[12] CARPENTER, J., 2011. May the best analyst win. *Science (New York, NY) 331*, 6018, 698.

[13] CARROLL, M., KOKINI, C., CHAMPNEY, R., SOTTILARE, R., and GOLDBERG, B., 2011. Modeling Trainee Affective and Cognitive State Using Low Cost Sensors. In *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*, Orlando, FL.

[14] CHAMPNEY, R.K. and STANNEY, K.M., 2007. Using emotions in usability SAGE Publications, 1044-1049.

[15] DASGUPTA, S. and LANGFORD, J., 2009. Active Learning Tutorial, ICML 2009.

[16] DORNEICH, M.C., VERVERS, P.M., MATHAN, S., and WHITLOW, S.D., 2007. *Defense Advanced Research Projects Agency (DARPA) Improving Warfighter Information Intake under Stress: Augmented Cognition - Phases 2, 3, and 4. - Final rept. Jun 2003-Jan 2007.* Honeywell, Inc., Honeywell Laboratories.

[17] FOWLER, P. and LEVINE, L., 1993. *A conceptual framework for software technology transition.* DTIC Document.

[18] GOLDBERG, B., BRAWNER, K., SOTTILARE, R., TARR, R., BILLINGS, D.R., and MALONE, N., 2012. Use of Evidence-based Strategies to Enhance the Extensibility of Adaptive Tutoring Technologies. In *The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC)* NTSA.

[19] GOLDBERG, B.S., SOTTILARE, R.A., BRAWNER, K.W., and HOLDEN, H.K., 2011. Predicting Learner Engagement during Well-Defined and Ill-Defined Computer-Based Intercultural Interactions. In *Proceedings of the 4th International Conference on Affective Computing and Intelligent Interaction (ACII 2011)*, *LNCS*, S.D. MELLO, A. GRAESSER, B. SCHULLER and J.-C. MARTIN Eds. Springer-Verlag, Berlin Heidelberg, 538-547.

[20] GONZALEZ, C., 2005. The relationship between task workload and cognitive abilities in dynamic decision making. *Human Factors 47*, 1, 92-101.

[21] HANLEY, J.A. and MCNEIL, B.J., 1983. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology 148*, 3, 839-843.

[22] HEWIG, J., HAGEMANN, D., SEIFERT, J., GOLLWITZER, M., NAUMANN, E., and BARTUSSEK, D., 2005. A revised film set for the induction of basic emotions. *Cognition and Emotion 19*, 7, 1095.

[23] JOHNSON, R.R., POPOVIC, D.P., OLMSTEAD, R.E., STIKIC, M., LEVENDOWSKI, D.J., and BERKA, C., 2011. Drowsiness/alertness algorithm development and validation using synchronized EEG and cognitive performance to individualize a generalized model. *Biological Psychology*.

[24] JONES, D., HALE, K., DECHMEROWSKI, S., and FOUAD, H., 2012. Creating Adaptive Emotional Experience During VE Training. In *The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC)* NTSA.

[25] KLEINSMITH, L.J. and KAPLAN, S., 1963. Paired-associate learning as a function of arousal and interpolated interval. *Journal of Experimental Psychology General 65*, 2, 190-193.

[26] KOEDINGER, K.R., BAKER, R., CUNNINGHAM, K., SKOGSHOLM, A., LEBER, B., and STAMPER, J., 2010. A data repository for the EDM community: The PSLC DataShop. *Handbook of Educational Data Mining*, 43-55.

[27] KOKINI, C., CARROLL, M., RAMIREZ-PADRON, R., HALE, K., SOTTILARE, R., and GOLDBERG, B., 2012. Quantification of trainee affective and cognitive state in real-time. In *The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC)* NTSA.

[28] LANDWEHR, N., HALL, M., and FRANK, E., 2005. Logistic model trees. *Machine Learning 59*, 1-2, 161-205.

[29] LANGFORD, J., LI, L., and STREHL, A., 2007. *Vowpal wabbit online learning.* Technical report.

[30] MCQUIGGAN, S., LEE, S., and LESTER, J., 2007. Early prediction of student frustration. *Affective Computing and Intelligent Interaction*, 698-709.

[31] SMALL, R.V., 1996. Dimensions of Interest and Boredom in Instructional Situations.

[32] SOTTILARE, R.A., BRAWNER, K.W., GOLDBERG, B.S., and HOLDEN, H.K., 2012. The Generalized Intelligent Framework for Tutoring (GIFT).

[33] WOOLF, B., BURLESON, W., ARROYO, I., DRAGON, T., COOPER, D., and PICARD, R., 2009. Affect-Aware Tutors: Recognizing and Responding to Student Affect. *International Journal of Learning Technology 4*, 3/4, 129--164.